A PSYCHO-ACOUSTICAL APPROACH TO QUANTIFY TIMBRE IN INSTRUMENTAL INDIAN MUSIC

Medha Basu*
Department of Physics, Jadavpur University, Kolkata, India
Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata, India
ORCID - https://orcid.org/0009-0007-4460-1245

Shankha Sanyal
Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata, India
School of Languages and Linguistics, Jadavpur University, Kolkata, India
ORCID – https://orcid.org/0000-0003-2879-0318

Sayan Nag
Adobe Research, India
ORCID - https://orcid.org/0000-0001-5652-125X

Kumardeb Banerjee
Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata, India

Dipak Ghosh
Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata, India

***E-mail of corresponding author:** medhabasu1996@gmail.com

Abstract:Indian Classical Music (ICM) is based on certain melodic frameworks, known as Ragas, which have the unique capability of evoking a variety of emotional experiences in the audience. The amount or character of emotional arousal varies from one Raga rendition to another and depends on the instrument in which it is played. Every instrument has its unique acoustic properties, which give rise to the signature style of rendition of that instrument. In this study, we have worked with two popular instruments of ICM, Sitar and Flute, belonging to the string and wind family respectively. Two traditional Ragas of ICM were chosen, - *Durga* and *Gunkali*, which convey contrasting emotions, happy and sad respectively (corroborated from a human response survey). Compositions based on each of these two ragas were chosen and recorded in both Sitar and Flute by two experienced musicians. Efforts were made to make the renditions of both the instruments as identical as possible with respect to melodic content and time duration. Small sections were prepared from the recordings, similar for both the instruments of a particular Raga, to carry out the next steps of timbral classification with help of linear and non-linear analysis techniques. For the linear analysis of the clips, spectral features (like spectral centroid, entropy, kurtosis, skewness etc.) of the acoustic waveforms were computed. In the first part of the study, we analyzed how accurately timbres of these two Indian instruments can be classified, solely depending upon spectral characteristics, when all other parameters like melodic content, time etc. are kept constant. From the complete set of the computed spectral features, statistical tests like one-way ANOVA were performed on the obtained data to establish the statistical significance of the obtained results at 95% confidence interval between the target classes. Mahalanabis Distance was computed to quantitatively understand the spatial distance between these two timbres based on the array of spectral features chosen to study. The dimensionality reduction technique of tSNE was applied in tandem with a KNN clustering approach to get the classification accuracy of timbres in case of melodies evoking different emotions. Next, the non-linear behavior of the raw acoustic waveforms was studied using a scaling analysis technique, Detrended Fluctuation Analysis (DFA), and Multifractal Detrended Fluctuation Analysis (MFDFA) with which, the long-range temporal correlations present in the acoustical signals were computed in terms of complexity determining parameter, Multifractal Width. Thus, from the results of both spectral and scaling analysis techniques used here, we envisaged to develop a novel approach for timbre comparisons and segregations in ICM, taking both linear and non-linear features into consideration.

Keywords: Indian Classical Music, Timber, Spectral features, DFA, tSNE, Mahalanobis Distance

I. INTRODUCTION

Music is a vast domain, including a huge variety of acoustic, emotion, genre-specific as well as location-specific characteristics. Indian Classical Music (ICM), the traditional music of the sub-continent of India, is home to innumerable instruments, world-renowned artists, different rendition styles etc. The most significant blueprint of ICM is the Raga system [1]. Every Raga is a unique melodic framework, consisting of a certain definite mode, tempo, and character of note-to-note transitions, which together has the unique power of evoking

emotions [2] [3] [4] [5] of various colours in humans. Various studies over years have confirmed the intertwined nature of music and emotion. Emotional elicitation and co-elicitations form one of the major characteristics of every musical rendition, be it vocal or instrumental. It is practically impossible to choose a music segment, which doesn't have any embedded emotion content. Rasa-theory [6], an age-old theory of emotion plays an important role here. Though the concept has undergone several modifications over centuries [7], it is still a widely accepted fact that music has the unique power of evoking different emotions of varied intensities in different individuals.

Timbre is a very closely associated term with acoustics and is defined as 'the quality of sound that is produced by a particular voice or musical instrument'. This is thus a qualitative concept with no specific scientific parameter defined for proper quantitative understanding. For two different instruments, even though we understand that the timbre or quality of the sound produced are different, it is not always associated with proper scientifically computed parameter-values. Specially in the domain of Indian instruments, no detailed work has been yet reported to explore the quantitative characterizations of timbre. We have tried to explore this idea in this work, with the aim of understanding timbre as a quantitative feature and what role emotions play here, with help of certain linear spectral features [8] and non-linear parameters. Two instruments have been chosen here, Flute [9] [10] and Sitar [11] [12], belonging to the wind and string families of instruments respectively. The flute is a member of a family of musical instruments in the woodwind group. Like all woodwinds, flutes are aerophones, producing sound with a vibrating column of air. Unlike woodwind instruments with reeds, a flute produces sound when the player's air flows across an opening. The other instrument, Sitar, belongs to the string family of instruments, and is played by plucking.

As already mentioned, every rendition has an associated emotional quotient with itself. In order to restrict the varying emotion of every music clip, it was necessary to keep the melodic content, tempo (beats per minute), and time duration identical for renditions by both the timbres. ICM is known for its unique power and character of improvisations, where the performer has the liberty to decorate his performance with different technical nuances and melodic ornamentations, within a definite framework of notes (*Ragas*). This makes every performance different, even for the same artist, thus making the application of feature analysis in this domain of music information retrieval extremely difficult. To tackle this problem as much as possible, and study feature analysis for Flute and Sitar, structured traditional compositions (popularly known as *bandish*) were taken as the input audio signals. As mentioned above, emotion arousal is also another very important aspect of ICM, and needs to be considered for any kind of music-related study. Keeping this in mind, *bandishes* corresponding to Ragas of two contrasting emotions (verified by human response survey), happy and sad, were taken for the study. Linear spectral features [13], like centroid [14] [15], entropy [16] [17], kurtosis [18], skewness [36], roll-off [41] and spread [37] have been calculated and compared for similar rendition by both the instruments.

Spectral centroid has been found to be an effective marker for emotional judgement for different timbres. The work by Wu et al. [30] concluded that a high-centroid instrument is likely to sound happier, and a low-centroid instrument is more likely to sound sadder. Spectral entropy is usually widely used in different applications of speech recognition. It is a measure of disorganization or uncertainty in a random variable, and assumes that speech spectrums are more organized than noise spectrums [31]. We have tried to apply this idea in case of instrumental music renditions, since we have worked with structured compositions, and thus extremely organized spectrums. It also captures the formants or peakiness of a distribution, and thus can be used to characterize timbres rendering the same melodic distribution here. Works have also been reported regarding application of different entropy kinds to distinguish different musical styles, textures, duration as well as pitch [32]. Entropy has also been used in a Hidden Markov Model (HMM) architecture for both speech and music segmentation [33] and the proposed method has shown a high percentage of performance accuracy. Works have been reported where kurtosis and skewness were compared for 18 musical instruments, showing clear distinction between string, woodwind and brass sounds [34]. Spectral spread is the measure of how evenly frequencies are distributed in a signal, i.e., signifying how much energy is spread across different frequencies. Here, since we have taken the same melodic content, it will be interesting to study how this parameter behaves for similar renditions in two different instruments. Spectral roll-off have been used for speech bandwidth classification and segmentation [35]. This is a measure of the skewness of the signal's frequency spectrum. Thus, how this skewness changes with timbre, when the melody content is kept constant is an interesting study. The spectral features were also statistically studied, using tools like one-way ANOVA [21], Mahalanobis Distance [22] [23] and tSNE [29] to study the effectivity of the chosen parameters in case of timbre classification and comparison.

Since audio signals are complex time-series with different fluctuations at different instants, it is also important to investigate the underlying non-linearity of the signals under study here. Multifractal Detrended Fluctuation Analysis (MFDFA) has been used to compute scaling parameters like

Multifractal Width and the Hurst exponent [19] [20] for identical clips of both the timbres. Various works have been reported where signal-complexity has been effective in classifying clips of different Indian Ragas as well as different musical instruments [38] [39]. Improvisation, being a very interesting part of music, has also been studied using these non-linear techniques [40].

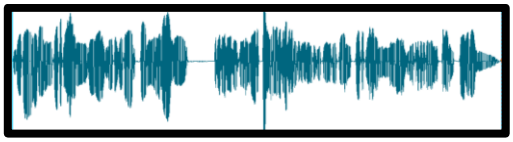
In a nutshell, in this pilot study, we have worked with two traditional *bandishes* of contrasting emotions, and have tried to establish a correlation between the different acoustical features with the timbral and emotional content of the compositions, with an aim to explore which feature performs timbre classification with maximum accuracy, and how the feature patterns change with change with the emotion conveyed by the melody chosen.

II. EXPERIMENTAL PROCEDURE

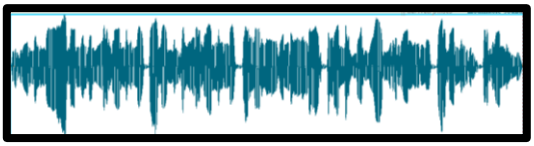
As mentioned in the previous section, it was necessary to choose the same structured compositions for both Flute and Sitar, so that the melodic, as well as emotional content of the music clips could be controlled. Two Ragas of contrasting emotions, happy and sad (as marked from an extensive emotion rating discussed in work [28]), were chosen (*Durga* and *Gunkali*) and the same traditional compositions for each of these Ragas were played and recorded in both the instruments. Two efficient and experienced students of ICM, both having received years of stringent training in Flute and Sitar respectively, were asked to record the compositions, without any kind of improvisation, and following the same tempo (bpm). Special attention was given so that no extra improvisation was recorded by any of the musicians, and the renditions by both the instruments were identical with respect to melodic content, tempo as well as time duration.

The detailed note framework along with the respective frequency spectrum of the two Ragas chosen for this study are-

- *Durga*: consisting notes - Sa Re Ma Pa Dha Sa



- *Gunkali*: consisting notes - Sa re Ma Pa dha Sa



In order to understand whether the same melody in both instruments evoke the emotions in similar manner, a detailed human emotion response survey was carried out with 50 participants. The compositions of each of the two Ragas (*Durga* and *Gunkali*) chosen for the study were used as the input stimuli for the survey, and participants (without prior familiarity to Indian Classical Music) were asked to mark the intensity elicitation corresponding to 4 emotions (Happiness, Sadness, Calmness, and Anxiety) for each of the two Ragas separately.

Next, from the recordings, identically sounding portions were selected with the help of audio editing software Wavepad [24] to prepare the clips for the study. For the analysis part of the experiment, two target classes were separately designed with the same audio recordings, one for linear spectral feature analysis and another for non-linear feature study. Differences between spectral features of both the timbres (with identical melodic sections) were initially verified with statistical tests like one-way ANOVA. Mahalanobis Distance was next calculated between the same set of feature values of Flute and Sitar, to study the spatial difference of the two chosen timbres corresponding to each of the chosen spectral parameters for happy and sad melodic clips in a more quantitative manner. Next, using the dimensionality reduction technique, tSNE, along with KNN clustering, the timbre classification accuracy of each of the two categories of happy and sad clips was studied, considering all the spectral features under a single feature space. Since audio signals are complex time series, characterized by non-linear frequency fluctuations, robust non-linear technique like Multifractal Detrended Fluctuation Analysis (MFDFA) along with Hurst exponent were used to carry out similar studies with the non-linear parameters. Each of the applied techniques and computed parameters are discussed briefly in the Methodology section.

III. METHODOLOGIES

A. Preparation of audio clip

As already discussed, two Ragas chosen for this study were *Durga* and *Gunkali*. At first, the audio signals for both Flute and Sitar were recorded in a soundproof studio, using the software Audacity [42]. The compositions chosen for each Raga consisted of two stanzas or proses, the first one called *Sthai* and the second one called *Antara*. For the Raga *Durga*, recording of the first prose (*Sthai*) in both Flute and Sitar was approximately 32 seconds each, and that of the second prose (*Antara*) was approximately 24 seconds each. Tempo maintained by both instruments was about 76-78 bpm (beats per minute). For the Raga *Gunkali*, recording of the first prose (*Sthai*) in both Flute and Sitar was approximately 33 seconds each, and that of the second prose (*Antara*) was approximately 38 seconds each. Tempo maintained in this Raga was about 80 bpm (beats per minute). These recordings were used for the next step of 'Human Response Survey'.

B. Human response Survey

Using the prepared recordings, a human emotion response survey was performed, to understand and compare levels of emotion elicitation from similar renditions of Sitar and Flute. Each form was supplied with 2 target emotions (happiness, sadness) and 2 distractor emotions (calmness, anxiety) with an intensity marking 5-point Likert scale (1- Very Low, 2- Low, 3- Moderate, 4- High and 5- Very High) for each emotion and response was collected from 50 participants, who were asked to mark the correct intensity of each of the emotions they felt after listening to each of the renditions of both the Ragas. A sample questionnaire is given below.

From the recorded response, corresponding to each Raaga clip separately, the average elicitation intensity of each of these 4 emotions were calculated using the following technique.

1). Average intensity –

Out of 4 emotions provided in the survey form (each with 5 intensities), one emotion was taken at a time. The various intensity levels for a particular emotion k were denoted by different values of variable i_s (i.e. very low is represented by i_1=1, low: i_2=2, moderate: i_3=3, high: i_4=4, and very high: i_5 =5). The number of participants who marked any particular intensity (i_s) for that particular emotion (k) was noted as n_{ik}. This number (n_{ik}) was then multiplied by the weight of the corresponding intensity (i_s). The total intensity value of a particular emotion (k) was obtained by doing a summation over these calculated weighted values for each intensity. Finally, dividing this total intensity by the total number of responses (N) (combining all emotions) recorded for the clip under consideration, we get the weighted average intensity $(AI)_k$ for the particular emotion (k) for that selected clip. In essence, following mathematical formula was used to calculate weighted average intensity value for each of the 8 emotions provided in the form:

$$(AI)_k = \sum_i (i_s n_{ik}) / N$$

Where,

- i_s = Intensity rating for emotion 'k',
  i_s € 1,2,3,4,5
- k = h, a, s, c, which signify each of the emotions under survey:
  (h: happiness, a: anxiety, s: sadness, c: calmness)
  $(AI)_k$ = weighted average intensity of emotion 'k' (for any particular clip)
- n_{ik} = no. of participants who marked for any particular intensity for emotion k
- N = total no. of participants who responded for the particular clip

Average intensity values for all the 4 emotions were calculated individually by the same method as explained in **Eqn. (1)** and was followed for all the clips of both *Durga* and *Gunkali*.

Comparisons revealed interesting information regarding happiness and sadness evoking quotient of Flute and Sitar.

This method was carried out before feature analysis in order to understand how feature characteristics of timbres depend on the emotion content of the melody being played by the instrument.

	Very Low - 1	Low - 2	Moderate - 3	High - 4	Very High - 5
Happiness					
Anxiety					
Sadness					
Calmness					

C. Linear Spectral Feature Analysis

Identical portions for both the timbres were next prepared with the help of Wavepad and Wavesurfer, which were used for the purpose of linear and non-linear feature comparisons. 7 such clips were prepared, separately for both the Ragas recordings (*Durga* and *Gunkali* evoking happy and sad emotions respectively). The chosen spectral features for this study are centroid, spread, skewness, kurtosis, entropy, and roll-off point. Each feature is elaborately discussed below.

1) Spectral centroid

Spectral centroid explains the frequency-weighted sum normalized by the unweighted sum. The spectral centroid represents the "centre of gravity" of the spectrum and is used as an indication of *brightness* and is commonly used in music analysis and genre classification.

$$\mu_1 = \frac{\sum_{k=b_1}^{b_2} f_k s_k}{\sum_{k=b_1}^{b_2} s_k} \quad \ldots (2)$$

where,

fk is the frequency in Hz corresponding to bin *k*, *sk* is the spectral value at bin *k*, *b*1 and *b*2 are the band edges.

2). Spectral Spread

Spectral spread is the standard deviation around the spectral centroid, representing the "instantaneous bandwidth" of the spectrum and is used as an indication of the dominance of a tone.

$$\mu_2 = \sqrt{\frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^2 s_k}{\sum_{k=b_1}^{b_2} s_k}} \quad \text{… (3)}$$

where,

fk is the frequency in Hz corresponding to bin *k*, *sk* is the spectral value at bin *k*, *b*1 and *b*2 are the band edges, and μ1 is the spectral centroid.

3). Spectral Skewness

Spectral skewness is computed from the third order moment, measuring symmetry around the centroid. In phonetics, spectral skewness is often used with other spectral moments to distinguish the place of articulation. For harmonic signals, it indicates the relative strength of higher and lower harmonics.

$$\mu_3 = \frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^3 s_k}{(\mu_2)^3 \sum_{k=b_1}^{b_2} s_k} \quad \text{… (4)}$$

where,

fk is the frequency in Hz corresponding to bin *k*, *sk* is the spectral value at bin *k*, *b*1 and *b*2 are the band edges, and μ1 is the spectral centroid, and μ2 is the spectral spread.

4). Spectral Kurtosis

Spectral kurtosis is computed from the fourth order moment, measuring the flatness, or non-Gaussianity, of the spectrum around its centroid. Conversely, it is used to indicate the peakiness of a spectrum.

$$\mu_4 = \frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^4 s_k}{(\mu_2)^4 \sum_{k=b_1}^{b_2} s_k} \quad \text{… (5)}$$

where,

fk is the frequency in Hz corresponding to bin *k*, *sk* is the spectral value at bin *k*, *b*1 and *b*2 are the band edges, and μ1 is the spectral centroid, and μ2 is the spectral spread.

5). Spectral Entropy

Spectral entropy measures the peakiness of the spectrum, used in voiced/unvoiced decisions for automatic speech recognition. Because entropy is a measure of disorder, regions of voiced speech have lower entropy compared to regions of unvoiced speech.

$$\textbf{Entropy} = \frac{-\sum_{k=b_1}^{b_2} s_k \log(s_k)}{\log(b_2 - b_1)} \quad \text{… (6)}$$

where,

fk is the frequency in Hz corresponding to bin *k*, *sk* is the spectral value at bin *k*, *b*1 and *b*2 are the band edges.

6). Spectral Roll-off Point

The spectral roll-off point measures the bandwidth of the audio signal by determining the frequency bin under which a given percentage of the total energy exists, and is used to distinguish between voiced and unvoiced speech, speech/music discrimination, music genre classification, acoustic scene recognition, and music mood classification.

Roll-off point = i such that $\sum_{k=b_1}^{i} |s_k| = \kappa \sum_{k=b_1}^{b_2} s_k$ **… (7)**

where,

sk is the spectral value at bin *k*, *b*1 and *b*2 are the band edges, κ is the specified energy threshold, usually 95% or 85%, *i* is converted to Hz before it is returned by spectral Roll-off Point.

D. Statistical tests

Certain statistical tests were performed on the computed feature values, which are discussed below.

1). One-way ANOVA

ANOVA is the technique of analysis of variance, where the means or averages of the datasets under study are compared with a definite confidence level (95% in this study) and p value (= 0.05 here) to study how significantly different the target class array values are from one another. Since we have worked with two timbres here, ANOVA was performed between the obtained feature values of Flute and Sitar, separately for all the parameters mentioned above. This was carried out for both the happy and sad Raga renditions, i.e., *Durga* and *Gunkali* respectively, to study whether emotion content affects the role of the chosen spectral parameters in the analysis of timbre classification.

2). **Mahalanobis distance**

In the original variable space, the parameter Mahalanobis Distance gives the measure of the correlation or the divergence or distance between data of various characteristics. It is calculated using the inverse of the variance–covariance matrix of the data set under study. Mahalanobis proposed this measure in 1930, has since then it has proved to be a very important statistical parameter.

For two groups labeled as $G1$ and $G2$, to compare and classify the two groups, we are often interested studying a number of relevant characteristics (say p) in both of these groups. X is defined as a vector which contains all the measurements made of a particular entity and is thus p-dimensional. In order to compare and classify the two groups,

we are often interested in measuring and summarizing their differences. Then the difference between the groups can be considered in terms of the difference between the mean vectors of X, in each group reactive to the common within-group variation. This measure is called the Mahalanobis squared distance defined by

$$\Delta^2=(\mu 1-\mu 2)\ T\ \sum^{-1}(\mu 1-\mu 2) \quad \ldots (8)$$

where 'T' denotes matrix transpose, $\sum$ denotes the common (non-singular) covariance matrix of X in each group $G1$ and $G2$. Since $\sum$ is a non-singular covariance matrix, it is positive defined and thus Δ is metric. Mahalanobis $\Delta 2$ is mainly used in classification problems, where there are several groups and the investigation concerns the affinities between groups.

In this study, we have calculated M.D. between the feature values of identical melodic segments of Flute and Sitar, separately for each of the 11 features under study. The M.D. values obtained help us to have a good understanding regarding the different levels of timbre classification abilities of the different parameters, and how these values change with the change in emotion content of the clips.

3). tSNE

tSNE, a dimensionality reduction tool, converts a large set of data in a high-dimensional space to a low-dimensional space, typically two-dimensional, preserving all the properties of the data set [43]. This algorithm helps in easier visualization and understanding of a large data set. From the feature-space of eight spectral features computed, each was treated as a separate dimension. The eight-dimensional feature set for each clip was converted to a two-dimensional space after application of dimensionality-reduction technique of tSNE, which led to a single plot. tSNE was performed between the feature-space of Flute and Sitar, for both the happy and sad clips, i.e., *Durga* and *Gunkali* renditions respectively.

KNN clustering was applied to compute the timbre classification accuracy for clips belonging to happy and sad emotion category separately, i.e., how effectively Flute and Sitar could be segregated from the 8-dimensional space of linear spectral features for clips of two contrasting emotion-content.

For a deeper insight into this study of timbre classification, we move on to the next step of non-linear feature analysis on the same clips.

E. Non-linear analysis

Next, we proceeded to analyze the non-linear behavior of the selected audio waveforms initially with the help of the robust non-linear techniques. Each signal was divided into 7 segments, and the scaling parameters were computed, which signify the long-range temporal correlation, denoting the complexity of the selected audio signal using Matlab programming. The question might arise, that is it possible to characterize the whole audio signal with one scaling exponent? There is a good possibility that the signal follows different patterns of complexities in different time instants i.e., shows a multifractal nature. So, to take this point into account, the multifractal width was finally calculated on the same signals, using the Multifractal Detrended Fluctuation Analysis (MFDFA) technique. This width is calculated from the computed multifractal spectrum, which considers the multifractality or different values of complexities in the signal. The details of the non-linear analysis technique are discussed below.

1). Multifractal Detrended Fluctuation Analysis (MFDFA)

The technique proposed by Kantelhardt (Kantelhardt 2008) was applied to compute the non-linearity and complexity audio signals investigating the underlying long-range temporal correlations. The following steps are followed.

i. Conversion of the noise-like audio signal to a simpler one Y(i), by deducting the signal mean value from every instant of the signal

$$\mathbf{Y(i) = \sum(x_k - x')} \quad \ldots(9)$$

Where, x' : mean value of the signal

ii. Segmentation of the entire signal Y(i) into Ns parts, with a definite number of samples of sample size 's' and the signal length 'N', the segments are denoted as

$$\mathbf{Ns = int(N/s)} \quad \ldots(10)$$

iii. Computation of the local RMS variation for sample size 's',

$$\mathbf{F^2(s,v)} = \frac{1}{s}\sum_{i=1}^{s}\{Y[(v-1)s+i] - \mathbf{y_v(i)}\}^2 \quad \ldots(11)$$

iv. Computation of the q-order overall RMS variation for different scale sizes

$$\mathbf{F_q(s)} = \{\frac{1}{Ns}\sum_{v=1}^{Ns}[\mathbf{F2(s,v)}]^{\frac{q}{2}}\}^{\frac{1}{q}} \quad \ldots(12)$$

v. Computation of the scaling behavior of the fluctuation function $F_q(s)$ from the log-log plot of $F_q(s)$ vs. s for each value of q **(Fig. 2a and b)**,

$$F_q(s) \sim s^{h(q)} \quad \ldots(13)$$

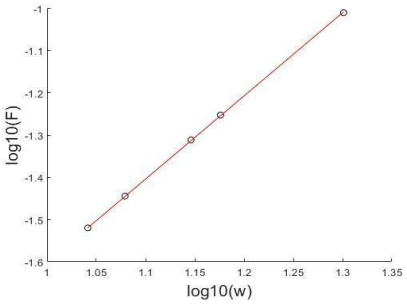


Fig 1. logF vs log(scale) plot for Hurst Exponent calculation

The slope of the best fit line obtained from the [logF vs log(scale)] plot $h(q)$, the generalized Hurst exponent which works as the autocorrelation or self-similarity parameter of the monofractal signal i.e., the nature of long-range correlation of the corresponding time series. A mono-fractal time series has a unique Hurst value of h(q) for all values of q.

This generalized Hurst exponent H(q) of MFDFA is related to the classical exponent τ(q) by the following equation:

$$\tau(q) = qh(q) - 1 \quad \ldots(14)$$

In case of a monofractal series with long range correlation, q order exponent τ(q) shows linear dependence with a single unique Hurst exponent h.

Majority of time series signals in nature are multifractal in nature and thus need multiple Hurst exponents at different instants to explain their non-linear behavior accurately. But, for multifractal signals, τ(q) shows non-linear relationship with q and the signal gets characterized by multiple Hurst exponents. The singularity spectrum f(α) of such a signal is then related to the Hurst exponent h(q) by the following equation:

$$\alpha = h(q) + qh'(q) \quad \ldots(15)$$

$$f(\alpha) = q[\alpha - h(q)] + 1 \quad \ldots(16)$$

Where, α denotes the singularity strength and f(α) denotes the dimension of the subset series, characterized by α. For q=2, α reduces to the conventional Hurst exponent, which has also been computed for this study. The range of exponents denotes the width of the multifractal spectrum. Applying the least square method, the multifractal spectrum can be characterized by fitting a quadratic function in the neighborhood of maximum α_0.

$$f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C \quad \ldots(17)$$

Where C is taken as an additive constant, $C = f(\alpha_0) = 1$ and B is a measure of the asymmetry or skewness of the spectrum. Value of B becomes zero for a perfectly symmetric signal i.e., equal left and right tails. A is a constant which determines the nature or orientation of the obtained spectrum. The width of the spectrum can be obtained by extrapolating the fitted quadratic curve to zero. Thereafter, detailed analysis of the plotted spectrums revealed various information regarding the extracted EEG signals.

Multifractal Width: This multifractal width W is defined as,

$$W = \alpha_1 - \alpha_2 \quad \ldots(18)$$

$$\text{With, } f(\alpha_1) = f(\alpha_2) = 0 \quad \ldots(19)$$

The multifractality of the spectrum is denoted by the width value of the spectrum, also called the multifractal width value. This value provides the measure of complexity of the signal. For a monofractal time series, this width reduces to zero since h(q) is independent of q in that case.

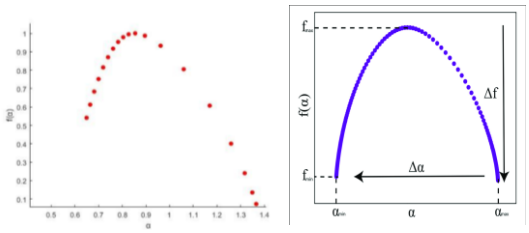


Fig 2. Multifractal spectrums

All the audio clips we have worked with in this study are complex time-series signals, and this technique of MFDFA was applied to compute the respective complexities.

In the computed spectrum, a special case was calculated for q = 2. In that case, the equation (15) gives a definite value of α = h(2). This scaling exponent works as the Hurst exponent with a definite value. This case has been reported for this work as well, and the Hurst exponents were computed for complex audio signals of all the renditions taken for the study.

IV. RESULTS AND DISCUSSIONS

A. Result from Audience Emotion Rating Survey

Previous studies have already revealed the emotion conveyed by '*Durga'* is prominent happiness and that of '*Gunkali'* is prominent sadness. Here, we have tried to understand what role timbre plays in emotion elicitation. Using similar structured renditions in both Flute and Sitar, response was

collected from 50 participants, who have no prior training or familiarity with ICM. The results for *Durga* (h1, h2) and *Gunkali* (s1, s2) renditions in Flute (F) and Sitar (S) are plotted together in the graphs below.

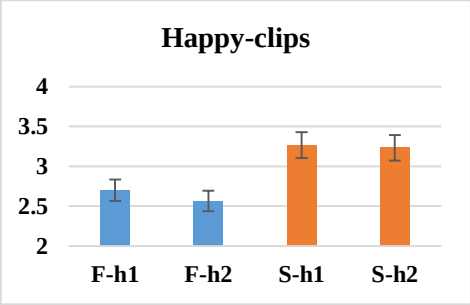


Fig. 3a. Happiness rating for Flute (F) and Sitar (S)

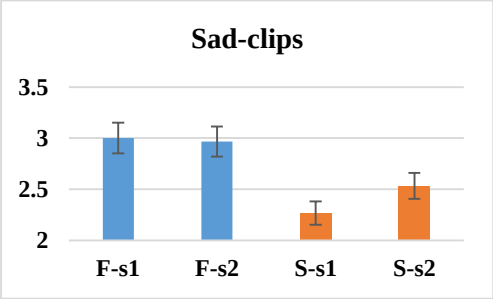


Fig. 3b. Sadness rating for Flute (F) and Sitar (S)

Some interesting trends can be observed from both the plots. In case of happy clips (*Durga*), Sitar shows higher average elicitation intensity than Flute and in case of sad clips (*Gunkali*), Flute shows higher average elicitation intensity than Sitar. Since the melodic components like tune, tempo, pitch etc. have been kept same for rendition by both the instruments, it can be concluded that this difference in emotion elicitation levels in both the instruments, is not due to any musical cause, rather due to the inner acoustic property of the instruments.

From the above discussed results, we aim to correlate the emotion eliciting property of these two instruments with the acoustic properties, in both linear and non-linear domains. The results for timbre classification using both linear and non-linear parameters for identical audio clips of Flute and Sitar, separately for these two Raga renditions are discussed below.

B. Timbre comparison from linear features

For each of the 6 features, comparative plots were drawn between the 7 chosen identical clips of Flute and Sitar, for both Ragas, *Durga* and *Gunkali*, reported as happy and sad emotion evoking melodies by the audience response survey. Comparative plots of timbre for both the Raga clips are shown and discussed in the plots below.

i. Spectral Centroid

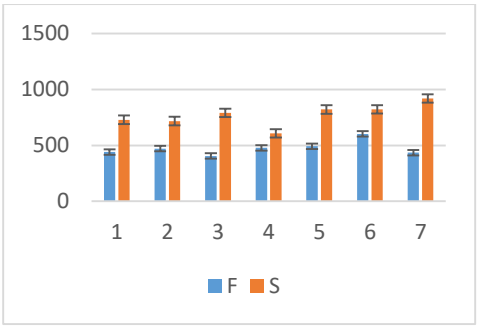


Fig. 4a.

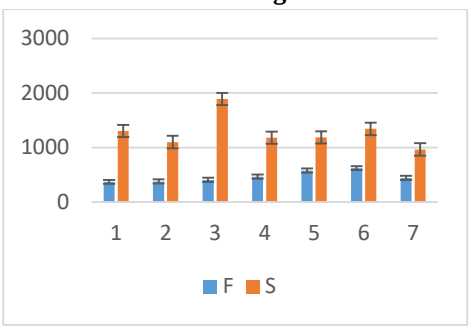


Fig. 4b.

ii. Spectral Spread

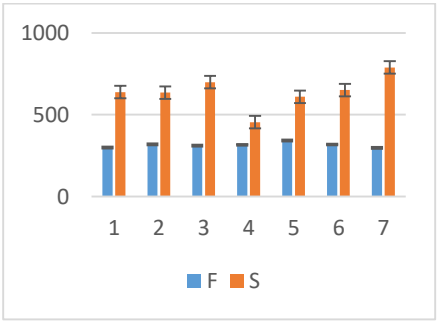


Fig. 5a.

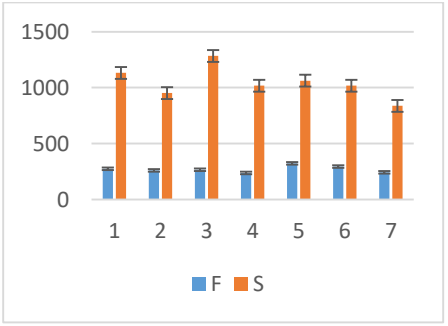


Fig. 5b.

iii. Spectral Skewness

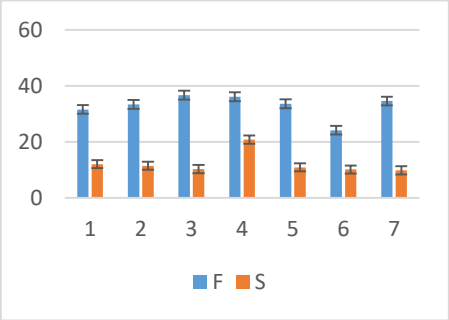


Fig. 6a.

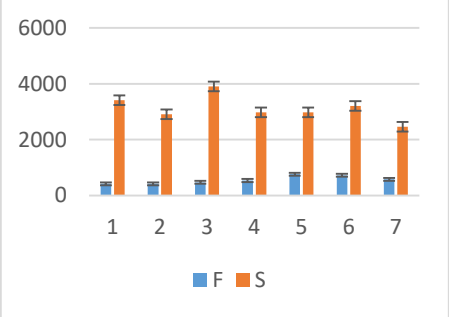


Fig. 6b.

iv. Spectral Kurtosis

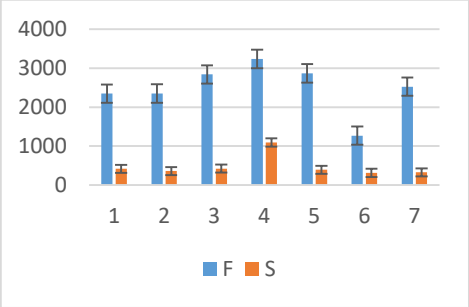


Fig.7a.

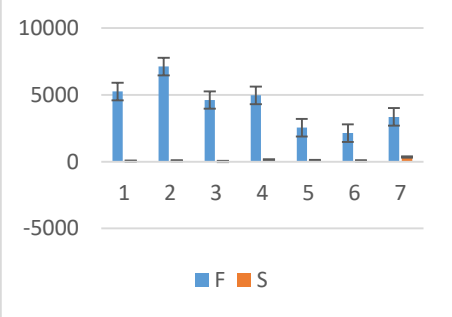


Fig. 7b.

v. Spectral Entropy

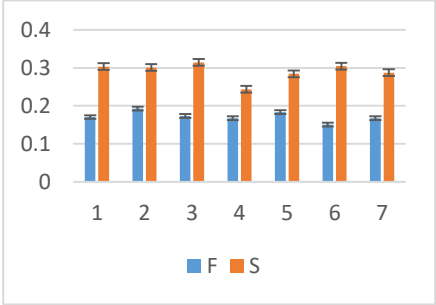


Fig. 8a.

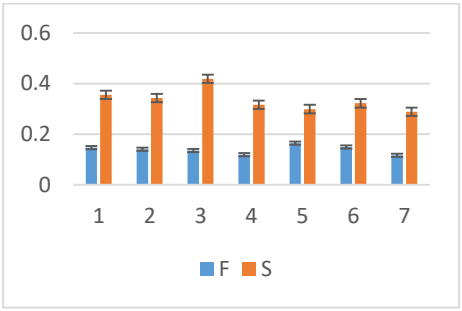


Fig. 8b.

vi. Spectral Roll-off

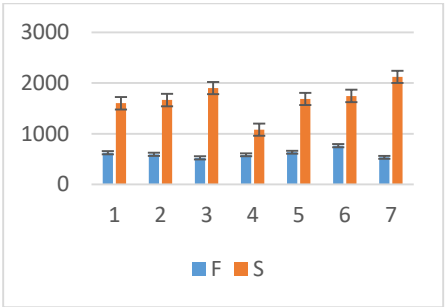


Fig. 9a.

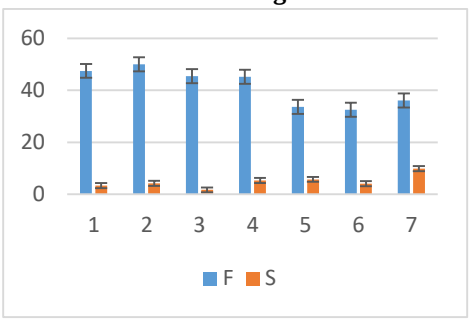


Fig. 9b.

Fig. Comparative plots for 4. Spectral Centroid
5. Spectral Spread
6. Spectral Skewness
7. Spectral Kurtosis
8. Spectral Entropy
9. Spectral Roll-off
for Flute (F) and Sitar (S) renditions for a). happy (Raga *Durga*) and b). sad (Raga *Gunkali*) clips,

From the plots above, certain interesting trends can be seen in acoustic properties of Flute and Sitar, corresponding to happy and sad emotion evoking clips.

Emotion independent feature behavior in seen in centroid, spread, entropy, and kurtosis (higher for Flute in centroid, spread, and entropy and for Sitar in kurtosis). Emotion dependent opposite feature behavior is seen in skewness, and roll-off (higher for Flute in happy clip and for Sitar in sad clip for both the features). The interesting dependency of timbre classification on emotion of the rendered clips are well-understood from the results discussed above.

Now, the results are statistically analyzed using the tools of ANOVA, Mahalanobis Distance, and tSNE. The results are discussed below.

i. ANOVA

As mentioned above, the spectral features mentioned in the previous section were computed for both happy and sad clips of Flute and Sitar. First, the one-way ANOVA test was performed on each of the features separately, between clips of two timbres of the same emotion, revealing the statistical difference between two timbral renditions evoking same emotion.

The result for both happy and sad clips of Flute and Sitar is shown in the table 2 below.

TABLE 1. ANOVA result between spectral features for Flute and Sitar, for happy (Durga) and sad (Gunkali) clips

	Happy clips			Sad clips		
	F value	**Probability**	**Significance**	**F value**	**Probability**	**Significance**
Centroid	47.43	0.000017	1	44.87	.000022	1
Spread	202.22	< .00001	1	70.73	< .00001	1
Skewness	162.08	< .00001	1	91.58	< .00001	1
Kurtosis	40.42	.000036	1	60.67	< .00001	1
Entropy	121.66	< .00001	1	138.65	< .00001	1
Roll-off	203.71	< .00001	1	74.2544	< .00001	1

All the selected features show significance value of 1 when ANOVA was performed between the feature values for identical clips, i.e., the set of 6 spectral features can be a good feature space for distinguishing between Flute and Sitar.

We move to a more detailed statistical study with these features with the next section of Mahalanobis distance calculation.

ii. Mahalanobis Distance

To study the timbre classification from a more detailed and quantitative point of view, Mahalanobis Distance was calculated between the feature values of Flute and Sitar, for clips with identical musical components. The chart below shows the comparative plot of Mahalanobis Distance values between identical renditions of Flute and Sitar, separately for each of the features. Results for both happy (*Durga*) and sad (*Gunkali*) clips are reported here. These values will give a detailed idea about the role of different features for timbre classification, and the dependency of emotions on the whole study.

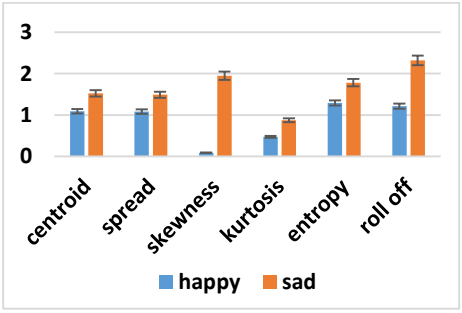


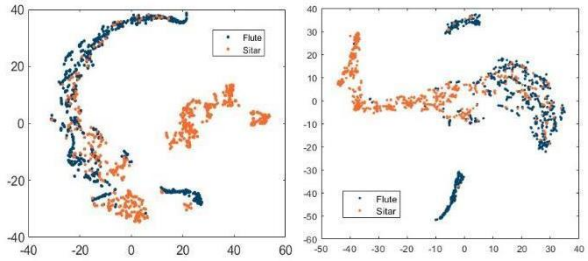
FIG. 10. Mahalanobis Distance between Flute and Sitar, separately for happy (Durga) and sad (Gunkali) clips

As can be seen from the plot, higher distance values are seen for sad (*Gunkali)* than happy (*Durga*) in all features which infers the better timbre classification for sad melodic renditions. High distance values refer to better classification. In case of *Durga* renditions, good timbre classification is shown by the features centroid, spread, entropy, and roll-off. Whereas, for *Gunkali* renditions, good timbre classification is shown by the features centroid, spread, skewness, entropy, and roll-off.

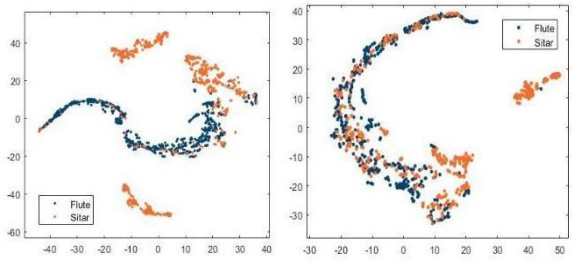
iii. tSNE and KNN clustering

Next, we have tried to understand this same phenomenon of timbral classification, using another statistical tool, the dimensionality reduction technique, tSNE. Unlike ANOVA and Mahalanobis Distance, where each of the computed features were studied separately, here all 8 features were treated as a single feature toolbox. Each of the 6 features (centroid, spread, skewness, kurtosis, entropy, and roll-off) were treated as a separate spectral dimension, and the final comparison through tSNE reduced this 6-dimensional space of the acoustic domain to a 2-dimensional space. The application of KNN clustering helped provide an idea about the accuracy index in timbre classification, for the identical clips of Flute and Sitar. The figures below show the timbre classification index for the two contrasting emotions, happy and sad, separately for the 7 audio portions prepared from the recordings, identical for both Flute and Sitar.

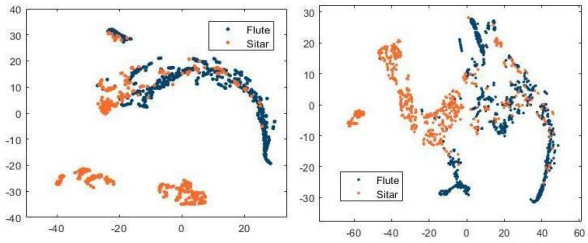
tSNE figures for happy clips along with the timbre classification accuracy index (A_i):



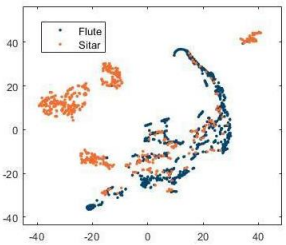
A_i = 89.71% **A_i = 88.98%**



A_i = 93.51% **A_i = 71.9%**

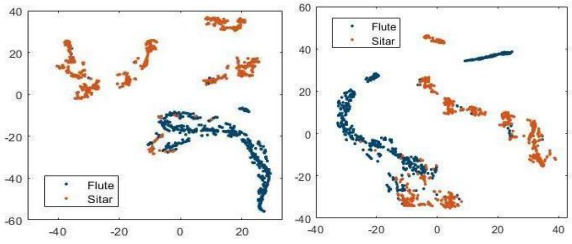


A_i = 87.17% **A_i = 88.92%**

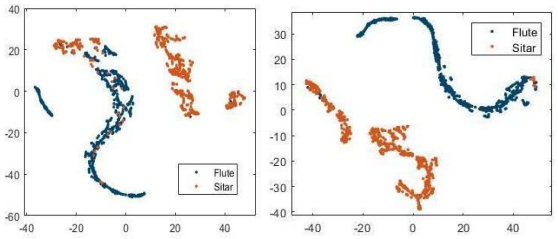


A_i = 91.11%

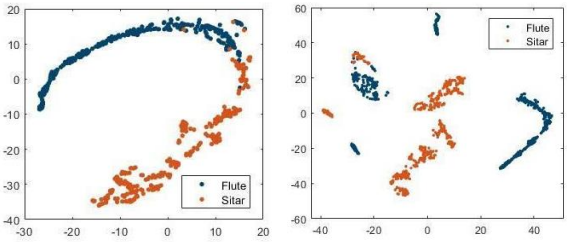
tSNE figures for sad clips along with the timbre classification accuracy index:



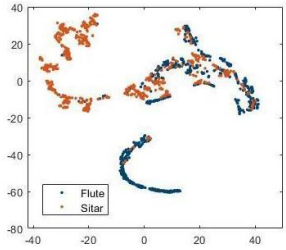
A_i = 98.59% **A_i = 98.3%**



A_i = 98.24% **A_i = 96.29%**



A_i = 94.13% **A_i = 95.45%**



A_i = 88.97%

The blue plots in the figures correspond to the feature points for Flute and the orange ones to the feature points for Sitar. The index values reveal that happy clips show average timbre classification accuracy of 87.33%, and sad clips show accuracy of 95.71%. Thus, it can be stated that sad clips show better timbre classification between Flute and Sitar, as was seen from the test of Mahalanobis Distance, when the mentioned spectral features are used as input dimensions for tSNE.

Next, we discuss the study of this timbre classification from the perspective of non-linear studies.

3. Timbre comparison from non-linear features:

We proceed to the non-linear method of timbre classification on the same audio clips, using parameter Multifractal Width (MW), which is a measure of multifractality or complexity of the audio signal (as discussed in the Methodology section). Since, like most other time-series signals, audio signal generated from music renditions are also characterized by non-linear and non-stationary fluctuations. Besides spectral feature analysis, study of these underlying non-linearity reveals more detailed information. The following graph explains the non-linear behavior of flute and sitar for both happy and sad clips with help of these two parameters. The reported values of MW (average of the computed values for 7 clips prepared, identical for Flute and Sitar) are given in

the table below (Table 2a), and the corresponding graph is plotted next.

TABLE 2a. Multifractal width values for Flute and Sitar, separately for happy *(Durga) and sad (Gunkali)*

	HAPPY CLIPS		SAD CLIPS	
	FLUTE	**SITAR**	**FLUTE**	**SITAR**
MW	0.92 ± 0.03	1.43 ± 0.06	1.03 ± 0.04	1.69 ± 0.08

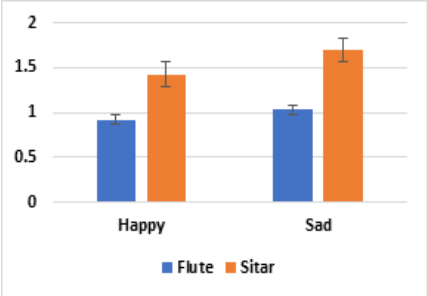


FIG. 11a. MW comparison

Fig. 11a. shows the multifractality trend for Flute and Sitar, which reveals high complexity i.e., higher MW value, for Sitar, in case of both happy (Durga) and sad (Gunkali) renditions.

Next, the exponent value for the case of q=2, was computed, and the Hurst exponents for the same audio clips were reported. The Hurst exponents (average of the computed values for 7 clips prepared, identical for Flute and Sitar) are given in the table below (Table 2b), and the corresponding graph (Fig 11b.) is plotted next.

TABLE 2b. Hurst exponent for Flute and Sitar, separately for happy *(Durga) and sad (Gunkali)*

	HAPPY CLIPS		SAD CLIPS	
	FLUTE	**SITAR**	**FLUTE**	**SITAR**
Hurst exponent	1.96 ± 0.06	1.73 ± 0.03	1.98 ± 0.08	1.36 ± 0.04

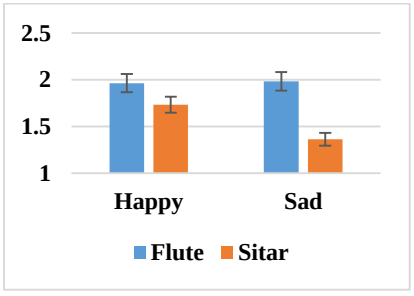


FIG. 11b. Hurst exponent comparison

Fig. 11b. shows the Hurst exponent (q=2) trend for Flute and Sitar, which reveals high high value for Flute, in case of both happy (Durga) and sad (Gunkali) renditions.

V. CONCLUSIONS

Certain observations can be made from the results discussed above.

- From the analysis of Audience Emotion Marking Survey, it can be distinctly seen that happiness elicitation is more pronounced in Sitar, and sadness in Flute, when similar renditions are presented in both the instruments.

- From the study of the linear spectral features of the selected clips, certain observations can be made. The results are discussed below.
  i. Emotion independent feature behavior is seen in the features centroid, spread, kurtosis, and entropy. Higher values were shown by Sitar in the features centroid, entropy, and spread, both for happy and sad clips. Higher values were shown by Flute in feature kurtosis.
  ii. Emotion dependent feature behavior, i.e., opposite trend for happy and sad music is seen in skewness, and roll-off. For skewness, higher values of Flute were shown for happy clips, and of Sitar for sad clips. For roll-off, higher value of Sitar was seen in happy clips, and of Flute in sad clips.

- Application of statistical techniques on the computed spectral features revealed the following information regarding timbre classification for identical melodic tunes.
  i. One-way ANOVA revealed the significant difference between all 6 spectral features considered for this study i.e., Centroid, Spread, Skewness, Kurtosis, Entropy, and Roll-off of the two timbres, equally prominent for both happy and sad clips.
  ii. Mahalanobis Distance shows a more detailed and quantitative picture of timbre classification that ANOVA. For the sad melodic segment, i.e., *Gunkali* rendition, better classification i.e., higher value of the parameter was observed between features of Flute and Sitar.
  iii. tSNE showed the timbral classification using all the 6 spectral features under a single feature space, revealing a better classification accuracy for sad clips.

- On the prepared clips, non-linear techniques of analysis applied for timbre classification led to the following conclusion.
    i. The non-linear parameter multifractal width shows higher values for Sitar, in case of both happy and sad clips. Thus, it can be said that the complexity of the instrument Sitar is always higher, irrespective of the emotional content of the renditions. For the special case of value 2 of q, we get the Hurst exponent, for which higher value is seen in Flute, for both happy and sad renditions.

The present pilot study revisits the concept of emotional arousal from a particular clip and emphasizes on the need for timbral considerations while quantifying the emotional elicitation. Here, in the first step we find higher happiness eliciting the capacity of Sitar and higher sadness eliciting the power of Flute. Both Spectral feature studies and non-linear analysis reveal interesting correlation-trends. For happy clips rendition in Sitar, a lower value of skewness and higher value of roll-off is reported. Thus, correlating both the studies, it can be stated that higher happiness elicitation power of Sitar is the psychological manifestation of lower skewness and higher roll-off levels in the acoustic domain. Similarly, results reveal that higher sadness elicitation power of Flute is the psychological manifestation of higher skewness and lower roll-off levels in the acoustic domain. We also derive some novel information regarding the emotion-independent spectral properties of centroid, spread, kurtosis and entropy, which show similar trends for both emotion-clips, and can be considered to be the spectral blueprints of Flute and Sitar. Non-linear studies add a unique dimension to the whole study, revealing higher complexity of Sitar, irrespective of the emotion conveyed through the rendition.

We plan to extend this pilot study to a larger dataset of instruments (timbres) and emotion domain (both primary and secondary emotions [26] [27]), with the aim of creating a correlation paradigm of emotions with different renditions of Indian instruments.

ACKNOWLEDGMENTS

MB (Ref No D-7/Sc/231/21) acknowledges the Department of Science and Technology (DST), Govt. of India, for providing the DST Inspire Fellowship for this study.

AUTHOR DECLARATIONS

Conflict of Interest

No conflict of interest was reported between the authors of the work.

Ethics Approval

Ethics approval was obtained for this work according to the ethical guidelines of the Ethical Committee of Jadavpur University (Approval No. 3/2013). All responses reported from the human response survey here have been used with complete consent of the participants.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Roy, Samarjit, Sudipta Chakrabarty, and Debashis De. "Time-based raga recommendation and information retrieval of musical patterns in Indian classical music using neural networks." *IAES International Journal of Artificial Intelligence* 6.1 (2017): 33.
2. 2Coan, James A., and John JB Allen, eds. *Handbook of emotion elicitation and assessment*. Oxford university press, 2007.
3. Amodio, David M., Leah R. Zinner, and Eddie Harmon-Jones. "Social psychological methods of emotion elicitation." *Handbook of emotion elicitation and assessment* 91 (2007): 91-105.
4. Fuentes-Sánchez, Nieves, et al. "Emotion elicitation during music listening: Subjective self-reports, facial expression, and autonomic reactivity." *Psychophysiology* 58.9 (2021): e13884.
5. Hunter, Patrick G., and E. Glenn Schellenberg. "Music and emotion." *Music perception* (2010): 129-164.
6. Chaudhury, Pravas Jivan. "The theory of rasa." *The Journal of Aesthetics and Art Criticism* 24.1 (1965): 145-149.
7. Patankar, R. B. "Does the" Rasa" Theory Have Any Modern Relevance?." *Philosophy East and West* 30.3 (1980): 293-303.
8. Agostini, Giulio, Maurizio Longari, and Emanuele Pollastri. "Musical instrument timbres classification with spectral features." *EURASIP Journal on Advances in Signal Processing* 2003 (2003): 1-10.
9. Boehm, Ted, Theobald Boehm, and Dayton C. Miller. *The flute and flute playing*. Courier Corporation, 2011.
10. Galway, James. *Flute*. Macdonald, 1982.
11. Miner, Allyn. *Sitar and Sarod in the 18th and 19th Centuries*. Vol. 7. Motilal Banarsidass Publ., 2004.
12. Burridge, Robert, Jay Kappraff, and Christine Morshedi. "The sitar string, a vibrating string with a one-sided inelastic constraint." *SIAM Journal on Applied Mathematics* 42.6 (1982): 1231-1251.
13. Peeters, Geoffroy. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project." *CUIDADO Ist Project Report* 54.0 (2004): 1-25.
14. Schubert, Emery, Joe Wolfe, and Alex Tarnopolsky. "Spectral centroid and timbre in complex, multiple instrumental textures." *Proceedings of the international conference on music perception and cognition, North Western University, Illinois*. sn, 2004.
15. Kua, Jia Min Karen, et al. "Investigation of spectral centroid magnitude and frequency for speaker recognition." *Odyssey*. 2010.

16. Toh, Aik Ming, Roberto Togneri, and Sven Nordholm. "Spectral entropy as speech features for speech 35recognition." *Proceedings of PEECS* 1 (2005): 92.
17. Misra, Hemant, et al. "Spectral entropy based feature fo robust ASR." *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 1. IEEE, 2004.
18. Antoni, Jérôme. "The spectral kurtosis: a useful tool for characterising non-stationary signals." *Mechanical systems and signal processing* 20.2 (2006): 282-307.
19. Qian, Bo, and Khaled Rasheed. "Hurst exponent and financial market predictability." *IASTED conference on Financial Engineering and Applications*. Cambridge, MA: Proceeding of the IASTED International Conference, 2004.
20. Mielniczuk, Jan, and Piotr Wojdyłło. "Estimation of Hurs exponent revisited." *Computational statistics & data analysis* 51.9 (2007): 4510-4525.
21. St, Lars, and Svante Wold. "Analysis of variance (ANOVA)." *Chemometrics and intelligent laboratory systems* 6.4 (1989): 259-272.
22. McLachlan, Goeffrey J. "Mahalanobi- distance." *Resonance* 4.6 (1999): 20-26.
23. De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L. Massart. "The mahalanobis distance." *Chemometrics and intelligent laboratory systems* 50.1 (2000): 1-18.
24. Saberzadeh, Venous. "Introduction of Wave Pad Software." *Interdisciplinary Journal of Virtual Learning in Medical Sciences* 2.3 (2011): 61-67.
25. Nag, Sayan, et al. "On the application of deep learning and multifractal techniques to classify emotions and instruments using Indian Classical Music." *Physica A: Statistical Mechanics and its Applications* 597 (2022): 127261.
26. Lopatovska, Irene. "Toward a model of emotions and mood in the online information search process." *Journal of the association for information science and technology* 65.9 (2014): 1775-1793.
27. Braniecka, Anna, et al. "Mixed emotions and coping: The benefits of secondary emotions." *PloS one* 9.8 (2014): e103940.
28. Sanyal, S., Banerjee, A., Basu, M., Nag, S., Ghosh, D., & Karmakar, S. (2020, December). Do musical notes correlate with emotions? A neuro-acoustical study with Indian classical music. In *Proceedings of meetings on acoustics* (Vol. 42, No. 1). AIP Publishing.
29. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(11).
30. Wu, B., Wun, S., Lee, C., & Horner, A. (2013, November). Spectral Correlates in Emotion Labeling of Sustained Musical Instrument Tones. In *ISMIR* (pp. 415-420).
31. Toh, A. M., Togneri, R., & Nordholm, S. (2005). Spectral entropy as speech features for speech recognition. *Proceedings of PEECS*, *1*, 92.
32. Margulis, E. H., & Beatty, A. P. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal*, *32*(4), 64-78.
33. Ajmera, J., McCowan, I., & Bourlard, H. (2003). Speech/music segmentation using entropy and dynamism features in a HMM classification framework. *Speech communication*, *40*(3), 351-363.
34. Dubnov, S., Tishby, N., & Cohen, D. (1997). Polyspectra as measures of sound texture and timbre. *Journal of New Music Research*, *26*(4), 277-314.
35. Kos, M., Kačič, Z., & Vlaj, D. (2013). Acoustic classification and segmentation using modified spectral roll-off and variance-based features. *Digital Signal Processing*, *23*(2), 659-674.
36. Kazazis, S., Depalle, P., & McAdams, S. (2022). Interval and Ratio Scaling of Spectral Audio Descriptors. *Frontiers in Psychology*, *13*, 835401.
37. Skaug, R., & Hjelmstad, J. F. (1985). *Spread spectrum in communication* (No. 12). IET.
38. Bhaduri, S., & Ghosh, D. (2016). Speech, music and multifractality. *Current Science*, 1817-1822.
39. Banerjee, A., Sanyal, S., Guhathakurata, T., Sengupta, R., & Ghosh, D. (2016). Categorization of stringed instruments with multifractal detrended fluctuation analysis. *arXiv preprint arXiv:1601.07709*.
40. Sanyal, S., Banerjee, A., Patranabis, A., Banerjee, K., Sengupta, R., & Ghosh, D. (2016). A study on improvisation in a musical performance using multifractal detrended cross correlation analysis. *Physica A: Statistical Mechanics and its Applications*, *462*, 67-83.
41. Sandhya, P., Spoorthy, V., Koolagudi, S. G., & Sobhana, N. V. (2020, December). Spectral features for emotional speaker recognition. In *2020 third international conference on advances in electronics, computers and communications (ICAECC)* (pp. 1-6). IEEE.
42. Audacity, T. (2017). Audacity. *The name audacity (R) is a registered trademark of dominic mazzoni retrieved from http://audacity. sourceforge. net*.
43. Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(Nov), 2579-2605.